NEW COMBINATORIAL DESIGNS AND THEIR APPLICATIONS TO GROUP TESTING*

K. A. Bush, W. T. Federer, H. Pesotan and D. Raghavarao

ABSTRACT

A class of designs with property C(t) are introduced for the first time, and their applications in group testing of samples are studied.

1. Introduction.

Let us consider the problem of classifying each of n given units into one of two disjoint categories called satisfactory and unsatisfactory (or, simply, good and bad or defective). The characteristic feature of group testing is that any number of units, say x, can be tested simultaneously, but the information obtained from a single test on x units, without any chance of error, is that either (i) all the x units are good, or (ii) at least one of the x units tested is bad, but it is unknown how many and which ones are bad. The problem is to devise a suitable method of classifying all the n units into good or bad categories with the least number of trials.

The first application of group testing in the literature was made by Dorfman [2] in pooling blood samples in order to classify each one of a large group of people as to whether or not they have a particular disease. Sobel and his co-workers [5], [8], [9], [10], [11] have devised various sequential procedures to classify the units and established the optimality of their results for large n. Lindström [3], [4] was interested in a slightly modified problem, in which each trial

* Research supported by Grant No. A8766 from the National Research Council of Canada and the NIH Grant No. GM-05900. The authors are alphabetically listed.

determines the exact number of defectives, and provided optimal procedures in a set-theoretic frame.

In Section 2 we introduce a new class of combinatorial designs, which we call t complete designs written as property $C(t)$ and study them in some detail. We then use them in group testing experiments. For terminology in combinatorics of design of experiments, we refer to Raghavarao [6]. By a design we refer to the set-symbol structure or its incidence matrix, depending upon the context.

2. A New Combinatorial Design.

Let S be a set of v symbols $1, 2, \cdots, v$ and let $B_1, B_2, \cdots, B_b$ be non-empty proper subsets of S for $i = 1, 2, \cdots, b$. The design D is the collection of subsets $B_1, B_2, \cdots, B_b$ along with the set of symbols S. We now define the following:

DEFINITION 2.1. *The design D is said to be t complete, written as the property $C(t)$, if for every t distinct symbols $\theta_1, \theta_2, \cdots, \theta_t \in S$,*

$$\bigcup_{j \in T} B_j = S - \{\theta_1, \theta_2, \cdots, \theta_t\} \tag{2.1}$$

where $T = \{j | \theta_i \notin B_j$ *for* $i = 1, 2, \cdots, t\}$.

The balanced incomplete block design (BIB design)

$$\begin{array}{c} (0, 1, 3); (1, 2, 4); (2, 3, 5); (3, 4, 6); (4, 5, 0); \\ (5, 6, 1); (6, 0, 2) \end{array} \tag{2.2}$$

with parameters

$$v = b = 7; \quad r = k = 3; \quad \lambda = 1 \tag{2.3}$$

has the property $C(2)$. For example, let us consider the symbols 0, 5.

The sets in which none of the symbols 0, 5 occur are (1, 2, 4), (3, 4, 6) and the union of these two sets is $\{1, 2, 3, 4, 6\}$. Similarly, if we consider the symbols 2, 6, the sets in which none of the symbols 2, 6 occur are (0, 1, 3), (4, 5, 0) whose union is $\{0, 1, 3, 4, 5\}$.

Trivially a $C(t)$ design exists for $1 \leq t \leq v$, with $b = v$. In fact, the design with $B_i = \{i\}$, for $i = 1, 2, \cdots, v$ is a $C(t)$ design for $1 \leq t \leq v$. We call such a design a trivial $C(t)$ design.

The class of designs to be considered to obtain $C(t)$ designs is not necessarily the BIB designs alone. Any kind of design may possess the property $C(t)$. We have the following theorem:

THEOREM 2.1. *A* BIB *design* v, r, k, b, λ *is* $C(t)$ *complete if*

$$r - t\lambda > 0 \, . \tag{2.4}$$

[Note: This theorem is still valid if a variance balanced BIB design is replaced by an equi-replicated pairwise balanced design of index λ; also if two combinatorial structures are isomorphic they have the same $C(t)$.

Proof. Let $\theta_1, \cdots, \theta_t$ be any set of t symbols and let u_j denote the frequency with which another symbol ϕ occurs in sets containing j of the t chosen symbols. Then

$$\sum_0^t u_j = r \quad \text{and} \quad \sum_1^t j u_j = t\lambda \, .$$

On subtracting

$$u_0 - \sum_2^t (j - 1) u_j = r - t\lambda \, .$$

When $r - t\lambda > 0$, each of the remaining $v - t$ symbols occurs at least $r - t\lambda$ times in sets disjoint from the selected t-plet.

This result is the best possible. The symmetric BIBD (7, 4, 2) with incidence matrix

$$\begin{matrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{matrix} \tag{2.5}$$

is not C(2) complete. If we delete row 1 of (2.5), we obtain an unequal block design which is also not C(2) complete. Then, we have

THEOREM 2.2. *If D_1 is any combinatorial structure which is t complete, and if D_2 is any design obtained from D_1 by deletion of a subset of rows, then D_2 is at most t complete. If D_3 is obtained from D_1 by deletion of a subset of columns, then D_3 is at least t complete.*

THEOREM 2.3. *The residual design of a symmetric BIB has the same C(t) as the parent design.*

Proof. Let $C(t|S)$ and $C(t|R)$ denote the completeness numbers for the symmetric parent and the residual design, respectively. Let $C^*(t|S)$ denote the completeness number for the design remaining after removal of one set of the parent design. Then $C(t|S) \geq C^*(t|S)$ by Theorem 2.2. Consider any t-plet in $C^*(t|S)$ composed of α rows from the residual and $t - \alpha$ other rows. Then if $\alpha < t$, there is no reduction in the value of $C^*(t|S)$ from this source as compared with $C(t|S)$ since the extra block in $C(t|S)$ would then automatically be eliminated. Therefore if there is a reduction in the value of t in $C^*(t|S)$, it must arise from the residual portion so that $C(t|S) \geq C(t|R)$. But evidently $C(t|R) \geq C(t|S)$, $t < v - r$, since one must consider t-plets from the residual portion in studying $C(t|S)$.

BIB designs with property $C(t)$ will have $b \geq v$. In a group-testing context we need designs with $b < v$ and some PBIB designs with 2 associates given to the Tables [1] can be verified to have property $C(t)$. The t-completeness property of PBIBD was given recently in a paper by Saha <u>et</u> <u>al</u>. [7].

In searching for designs with the $C(t)$ property in the known classes of designs, the following theorem will be helpful.

THEOREM 2.4. <u>If a design D with sets</u> $S_1, S_2, \cdots, S_b$ <u>and</u> S <u>as the set of symbols has the property</u> $C(t)$, <u>then the complementary design</u> D^* <u>formed from the sets</u> $S_1^*, S_2^*, \cdots, S_b^*$, <u>the number of times every t-plet of symbols</u> $(\theta_1, \theta_2, \cdots, \theta_t)$ <u>occurs, denoted by</u> $\lambda_{\theta_1\theta_2\cdots\theta_t}$ <u>is a positive integer, where</u> $S_i^* = S - S_i$, <u>for</u> $i = 1, 2, \cdots, b$.

<u>Proof</u>. When the design D has property $C(t)$, there exist sets, say, $S_{i_1}, S_{i_2}, \cdots, S_{i_x}$ where a given t-plet of symbols $(\theta_1, \theta_2, \cdots, \theta_t)$ does not occur while all the other symbols occur at least once. Then in D^*, the blocks $S_{i_1}^*, S_{i_2}^*, \cdots, S_{i_x}^*$ will each contain the symbols $\theta_1, \theta_2, \cdots, \theta_t$ and hence $\lambda_{\theta_1\theta_2\cdots\theta_t} = x\ (> 0)$.

The condition stated in the theorem is only necessary, but not sufficient. The BIB design with parameters $v = b = 7$, $r = k = 4$, $\lambda = 2$ has its complementary design in which every symbol occurs at least once satisfying the condition of the theorem but does not possess the $C(2)$ property as indicated after Theorem 2.1.

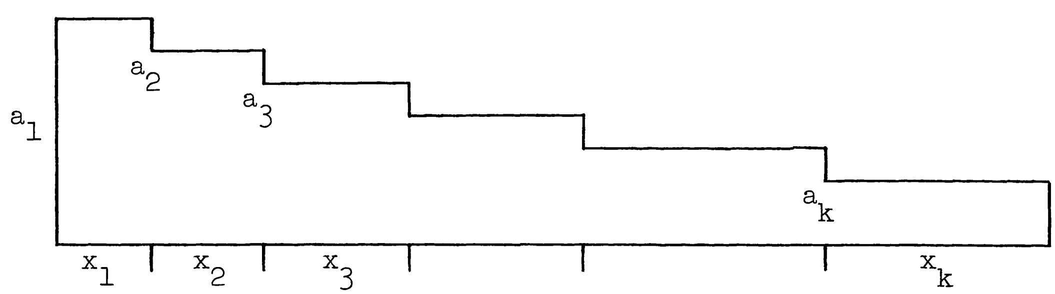
It is well known that $C[k, t, \delta, v]$ configurations (or t-designs) are also $C[k, t', \delta', v]$ configurations (or t'-designs) for $t' < t$. Analogous to this result we have the following:

THEOREM 2.5. _If a design D has the property $C(t)$, then it has the property $C(t - 1)$ for $2 \leq t \leq v$._

Proof. Let $(\theta_1, \theta_2, \cdots, \theta_t)$ be any t-plet. Among the sets where at least one of the symbols $\theta_1, \theta_2, \cdots, \theta_t$ occurs, for each θ_i $(i = 1, 2, \cdots, t)$ there exists at least one set in which θ_i occurs without $\theta_1, \theta_2, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_t$. Otherwise for the t-plet $(\theta_1, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_t, \phi)$ with $\phi = \theta_i$ the property $C(t)$ for the design D will be violated. Now the sets in which none of $(\theta_1, \theta_2, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_t)$ occurs is when all the other $v - t + 1$ symbols occur proving that D has property $C(t - 1)$.

3. Designs with $b < v$ Having Property $C(1)$.

Any v can be written in the form $a_1x_1 + a_2x_2 + \cdots + a_kx_k$ where $k = 1, 2, \cdots$ and a_i, x_i are positive integers. Without loss of generality, we assume $a_1 > a_2 > \cdots > a_k$. The numbers can then be written in the form of a staircase:



Now form b sets where $b = a_1 + x_1 + x_2 + \cdots + x_k$ by writing the symbols in the a_1 rows and the $x_1 + x_2 + \cdots + x_k$ columns. Such designs will have exactly 2 replications for each symbol and have various cardinalities for the sets constituting the design. A moment's consideration into the above construction indicates that such designs have the property C(1).

Let us illustrate our construction method for $v = 19$. Since $19 = 4 \times 4 + 3 \times 1$, we write the 19 symbols in the staircase array

$$\begin{array}{ccccc} 1 & 2 & 3 & 4 & \\ 5 & 6 & 7 & 8 & 19 \\ 9 & 10 & 11 & 12 & 18 \\ 13 & 14 & 15 & 16 & 17 . \end{array} \tag{3.1}$$

We now form 9 sets for the design by writing the sets formed from the rows and columns of the array (3.1) to get

$$\begin{array}{l} (1,\ 2,\ 3,\ 4);\ (5,\ 6,\ 7,\ 8,\ 19);\ (9,\ 10,\ 11,\ 12,\ 18); \\ (13,\ 14,\ 15,\ 16,\ 17);\ (1,\ 5,\ 9,\ 13);\ (2,\ 6,\ 10,\ 14); \\ (3,\ 7,\ 11,\ 15);\ (4,\ 8,\ 12,\ 16);\ (17,\ 18,\ 19). \end{array} \tag{3.2}$$

The design (3.2) can be easily verified to have the property C(1).

It is interesting to note that as the partitioning of v as $a_1x_1 + a_2x_2 + \cdots + a_kx_k$ is not unique and different partitionings give different numbers of sets, it is desirable to consider the partitioning for a given v which minimizes $b = a_1 + x_1 + x_2 + \cdots + x_k$.

We now study the existence of C(1) designs with $b < v$ and the asymptotic property for b/v as $v \to \infty$. These results are given in the following theorem.

THEOREM 3.1. <u>Designs with property</u> C(1) <u>and</u> $b < v$ <u>exist for all</u> $v \geq 6$. <u>Furthermore,</u>

$$\lim_{v \to \infty} \frac{b}{v} = 0 \ . \tag{3.3}$$

<u>Proof.</u> Let $m^2 < v \leq (m+1)^2$, for $m = 0, 1, \cdots$. We distinguish two cases: case (i) $v = m^2 + a$, where $1 \leq a \leq m$ and case (ii) $v = m^2 + m + b$, where $1 \leq b \leq m + 1$. In case (i) clearly $v = m \times m + a \times 1$ and from our earlier consideration a design with property C(1) can be constructed in $b = m + (m+1) = 2m + 1$ sets. In case (ii) we have $v = (m+1) \times m + b \times 1$ and we can construct a design with property C(1) in $b = (m+1) + (m+1) = 2(m+1)$ sets. Again $2m + 2 < (m+1)m + b$ where $1 \leq b \leq m+1$ for all $m \geq 3$. These considerations imply that $b < v$ for all $v \geq 6$. Now

$$\frac{b}{v} = \frac{2m + i}{v} = o\left(\frac{1}{m}\right) , \tag{3.4}$$

where $i = 1$ or 2 depending on whether v belongs to case (i) or (ii) and the assertion (3.3) follows.

The designs with property C(1) constructed by the above staircase method will not always give the smallest b and this follows from the following theorem:

THEOREM 3.2. <u>If</u> D_i $(i = 1, 2)$ <u>are designs with property</u> C(1) <u>on</u> v_i <u>symbols in</u> b_i <u>sets, then there exists a design</u> $\tilde{D}$ <u>with property</u> C(1) <u>on</u> $v_1 v_2$ <u>symbols in</u> $(b_1 + b_2)$ <u>sets.</u>

<u>Proof.</u> Let S_i be the set of v_i symbols and let $S = S_1 \times S_2$ where '$\times$' is the Cartesian product of sets. Let $B_{1i}, B_{2i}, \cdots, B_{b_i i}$ be the sets of the design D_i. Consider the $b_1 + b_2$ sets $S_1 \times B_{j2}$ and $B_{\ell 1} \times S_2$ for $j = 1, 2, \cdots, b_2$ and $\ell = 1, 2, \cdots, b_1$ constituting the design $\tilde{D}$ on $v_1 v_2$ symbols of $S_1 \times S_2$. It can easily be verified that $\tilde{D}$ has property C(1).

Using the methods of (3.1), we have a design with property C(1) on 8 symbols in 6 sets. From this design, using Theorem 3.2, we can construct a design with property C(1) on 64 symbols in 12 sets. The design given using (3.1) on 64 symbols with property C(1) has 16 sets, while Theorem 3.2 leads us to a design with only 12 sets. Consequently, for $64^2 = 4096$ symbols from Theorem 3.2, we can construct a design with 24 sets while the design constructable using (3.1) has 128 sets. Thus we achieve considerable reduction in the number of sets used in the design by using the method of Theorem 3.2.

THEOREM 3.3. *The largest* v *that can be used in a* C(1) *design with fixed* b *is* $\binom{b}{[b/2]}$ *where* $[b/2]$ *is the integral part of* $b/2$. *Such a design is the dual of the unreduced* BIB *design with parameters*

$$b\binom{b-1}{[b/2]-1} \quad [b/2] \quad \binom{b}{[b/2]} \quad \binom{b-2}{[b/2]-1}.$$

Proof. For a combinatorial structure to fail to be C(1) complete, it is clearly necessary that there must be two rows of the incidence matrix with inner product r_1 where r_1 is the number of ones in one of these two rows. Thus any design whose rows are all the $\binom{b}{r}$ r-tuples which can be formed from the b objects (or any subset of them) is C(1) complete. The largest number of rows constructable by these means occur when $r = [b/2]$, and this is best possible by Sperner's [12] theorem.

Sperner defined two subsets of the integers 1, 2, $\cdots$, b as independent if neither set is a subset of the other, and he proved that the maximum number of independent sets was

$$\binom{b}{[b/2]}.$$

Thus if we form this largest number of independent subsets of 1, 2, $\cdots$, b using them as the rows of the configuration, then the inner product of any two rows of the incidence matrix is smaller than $[b/2] = \min(r_1, r_2)$, and we have C(1) completeness.

In the reverse direction, we have: if

$$\binom{b-1}{[(b-1)/2]} < v \leq \binom{b}{[b/2]},$$

then the least number of blocks to produce a C(1) complete configuration is b.

4. Applications of Designs with property C(t) in Group Testing Experiments.

Let there be v units in the population and let it be known to the experimenter before hand that there are exactly t units in the population which are defective, while v - t units are good. Further, it is unknown to the experimenter which of the units are defective. Then one can make b tests (or runs) on the v units, where each test is made on the collection of the units belonging to the sets of a design D with property C(t). If the test gives a negative result, the units involved in the test are all good and if the test gives a positive result, at least one of the units involved in the test is bad. If x tests give negative results in each test and the remaining b - x tests give positive results, the C(t) property of the design guarantees that the units, included in the set union of the sets corresponding to the negative test results, are all good which will be v - t in number, while the other v - t are bad.

As an illustration, let us consider that there are 12 units among which we know that 2 are bad and 10 are good. The design SR41 of the Tables in [1] has property C(2). The test number and the units tested in each are as follows:

Test Number	Units Included in the Test
1	1, 2, 3, 4
2	7, 10, 5, 4
3	6, 11, 9, 4
4	1, 7, 6, 8
5	11, 5, 2, 8
6	10, 9, 3, 8
7	1, 11, 10, 12
8	9, 2, 7, 12
9	5, 3, 6, 12

The classification of items and the test numbers indicating negative results are as follows:

Test Number	Defective Items	Test Number	Defective Items
2, 3, 6, 9	1, 2	7, 8, 9	4, 8
2, 3, 5, 8	1, 3	4, 5, 7, 9	4, 9
5, 6, 8. 9	1, 4	4, 5, 8, 9	4, 10
3, 6, 8	1, 5	4, 6, 8, 9	4, 11
2, 5, 6, 8	1, 6	4, 5, 6	4, 12
3, 5, 6, 9	1, 7	1, 6, 7, 8	5, 6
2, 3, 8, 9	1, 8	1, 3, 6, 7	5, 7
2, 5, 9	1, 9	1, 3, 7, 8	5, 8
3, 5, 8, 9	1, 10	1, 4, 7	5, 9
2, 6, 8, 9	1, 11	1, 3, 4, 8	5, 10
2, 3, 5, 6	1, 12	1, 4, 6, 8	5, 11
2, 3, 4, 7	2, 3	1, 3, 4, 6	5, 12
4, 6, 7, 9	2, 4	1, 5, 6, 7	6, 7
3, 4, 6, 7	2, 5	1, 2, 7, 8	6, 8
2, 6, 7, 9	2, 6	1, 2, 5, 7	6, 9
3, 6, 7, 9	2, 7	1, 5, 8	6, 10
2, 3, 7, 9	2, 8	1, 2, 6, 8	6, 11
2, 4, 7, 9	2, 9	1, 2, 5, 6	6, 12
3, 4, 9	2, 10	1, 3, 7, 9	7, 8
2, 4, 6, 9	2, 11	1, 5, 7, 9	7, 9
2, 3, 4, 6	2, 12	1, 3, 5, 9	7, 10
4, 5, 7, 8	3, 4	1, 6, 9	7, 11
3, 4, 7, 8	3, 5	1, 3, 5, 6	7, 12
2, 5, 7, 8	3, 6	1, 2, 7, 9	8, 9
3, 5, 7	3, 7	1, 3, 8, 9	8, 10
2, 3, 7, 9	3, 8	1, 2, 8, 9	8, 11
2, 4, 5, 7	3, 9	1, 2, 3	8, 12
3, 4, 5, 8	3, 10	1, 4, 5, 9	9, 10
2, 4, 8	3, 11	1, 2, 4, 9	9, 11
2, 3, 4, 5	3, 12	1, 2, 4, 5	9, 12
4, 6, 7, 8	4, 5	1, 4, 8, 9	10, 11
5, 6, 7, 8	4, 6	1, 3, 4, 5	10, 12
5, 6, 7, 9	4, 7	1, 2, 4, 6	11, 12

In view of Theorem 3.1, if a large population has exactly 1 bad item, it can be detected in b tests, where b is only a very small fraction of v.

Acknowledgment.

The authors wish to thank Professor B. Leo Raktoe for many helpful discussions.

REFERENCES

[1] Clatworthy, W. H. (1973). Tables of two-associate-class partially balanced designs. *NBS Appl. Math. Series 63*

[2] Dorfman, B. (1943). The detection of defective members of large populations. *Ann. Math. Statist.* 14, 436-440.

[3] Lindstrom, B. (1964). On a combinatory detection problem. I. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, IX, 195-207.

[4] Lindstrom, B. (19). Determining subsets of unramified experiments. Unpublished paper.

[5] Kumar, S. and M. Sobel (1971). Finding a single defective in binomial group testing. *J. Amer. Statist. Assoc.* 66, 824-828.

[6] Raghavarao, D. (1971). *Constructions and Combinatorial Problems in Design of Experiments.* Wiley.

[7] Saha, G. M., H. Pesotan, and B. L. Raktoe (1982). Some results on t-complete designs. *Ars Combinatoria* 13, 195-201.

[8] Sobel, M. (1960). *Group-testing to Classify all Defectives in a Binomial Sample.* A contribution in Information and Decision Processes, R. E. Machol, ed., McGraw-Hill, 127-161.

[9] Sobel, M. (1967). Optimal group testing. *Proc. Colloquium on Information Theory Organized by the Bolyai Mathematical Society* (Hungary), 411-488.

[10] Sobel, M. and P. A. Groll (1959). Group testing to eliminate efficiently all defectives in a binomial sample. *Bell System Tech. J.* 38, 1179-1252.

[11] Sobel, M. and P. A. Groll (1966). Binomial group testing with an unknown proportion of defectives. *Technometrics* 8, 631-656.

[12] Sperner, E. (1928). Ein Satz uber Untermengen einer endliche Menge. *Math. Zeit.* 27, 544-548.

1. Washington State University
Pullman, Washington

2. Cornell University
Ithaca, New York

3. University of Guelph
   Guelph, Ontario

4. Temple University
   Philadelphia, Pennsylvania